



Data-Driven User Guidance in Multi-Attribute Data Exploration

Masterstudium:
Medizinische Informatik

Klaus Eckelt

Technische Universität Wien
Institute of Visual Computing & Human-Centered Technology
Arbeitsbereich: Research Division of Computer Graphics
Betreuer: Ao.Univ.Prof. Univ.-Doz. Dipl.-Ing. Dr.techn. Eduard Gröller
Mitbetreuung: Univ.-Prof. Dipl.-Ing. Dr.techn. Marc Streit

Motivation

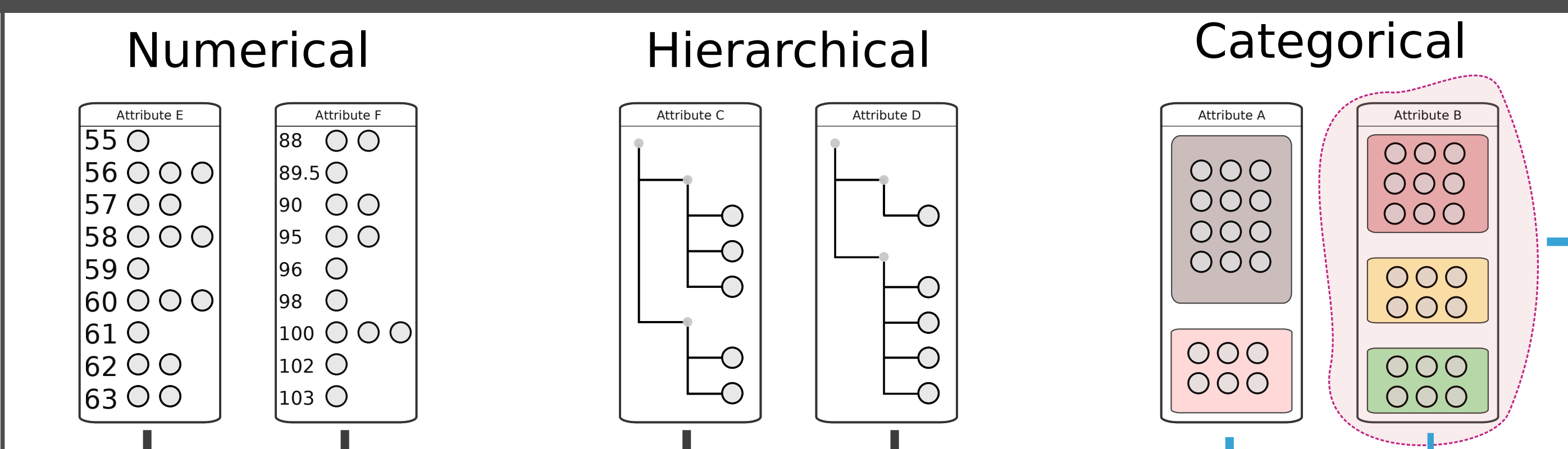
Data analysis requires computational power and human reasoning to cope with the vast volume or inherent complexity [1]. While humans can quickly identify relationships and patterns in a visualization, the computational power surpasses them in the calculation of defined measures across whole datasets. The measures can be used to guide users towards promising data subsets for visual exploration.

Objectives

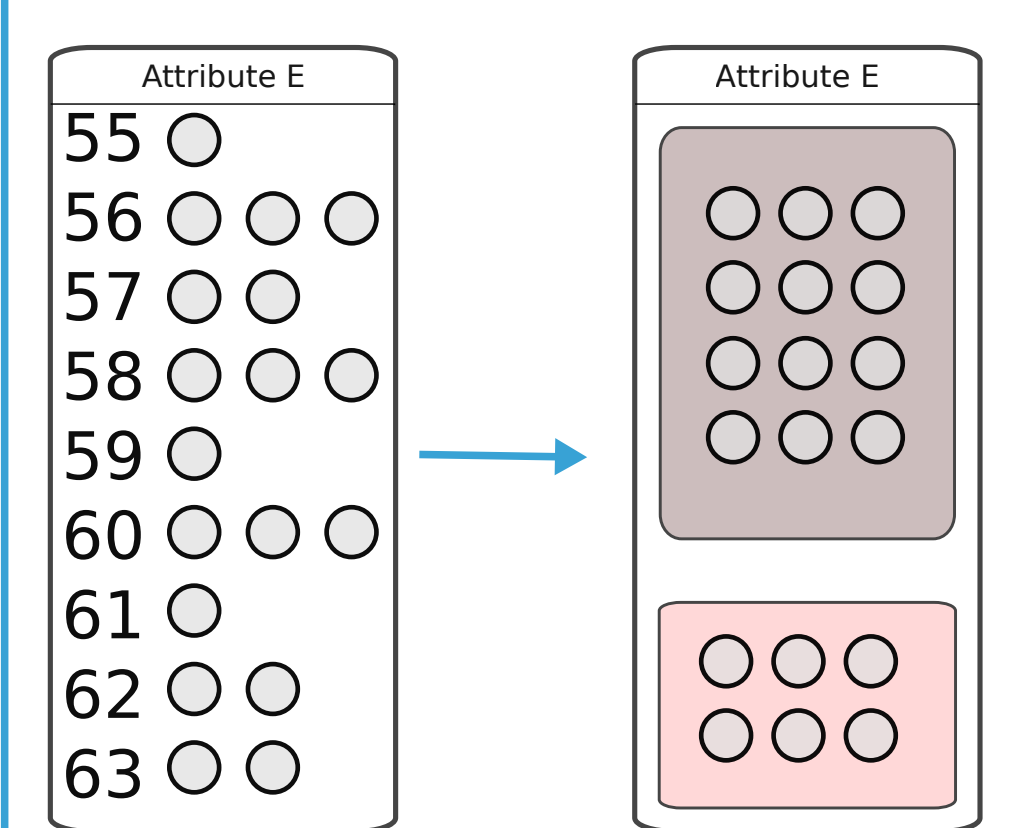
A touring system, to find attributes of interest for the users, that is:

- generalized for different data types.
- independent of the data domain.
- independent of the visual analysis tool.

Attributes

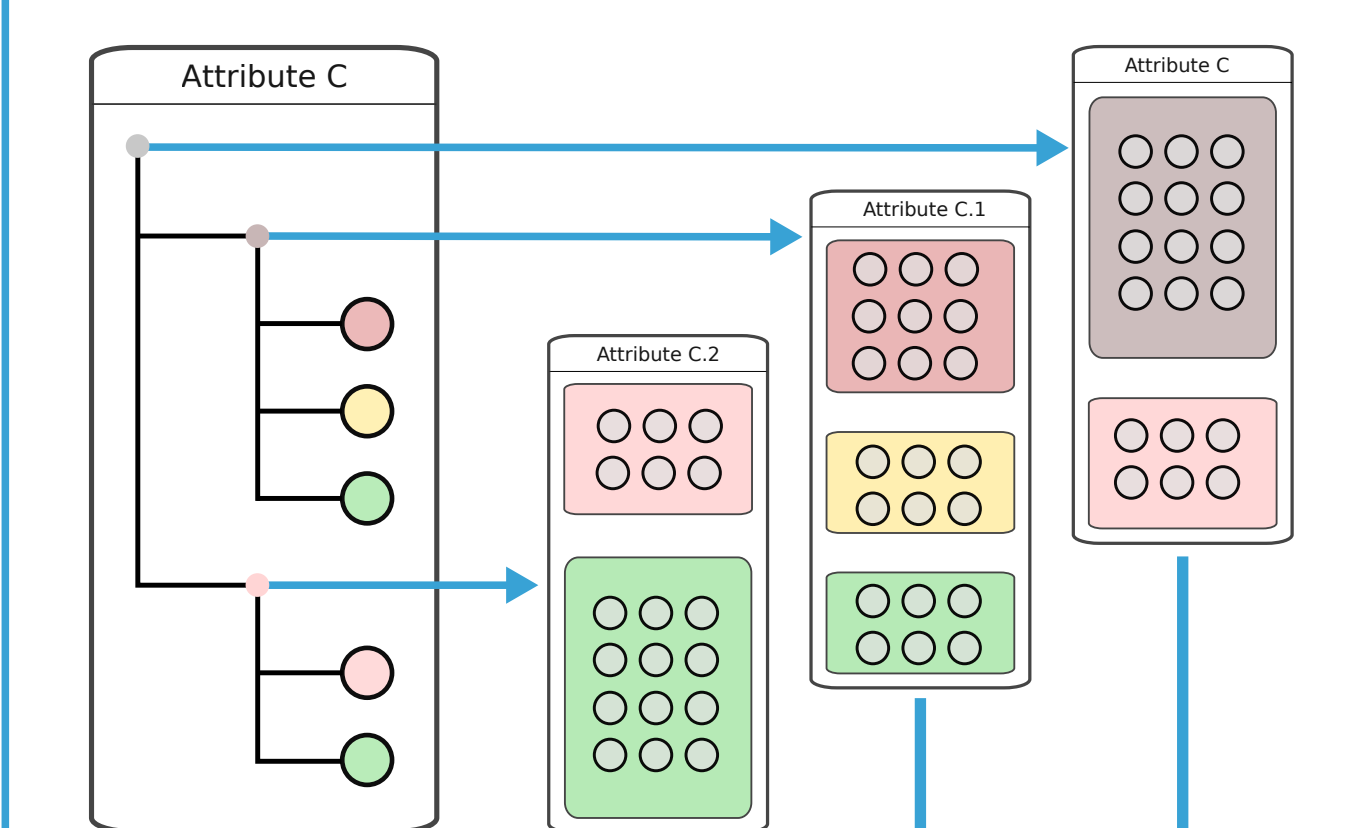


Discretization



The numerical values are split into two categories, such that one of them has maximum similarity. A separate category may be created for items with no recorded values.

Categorization

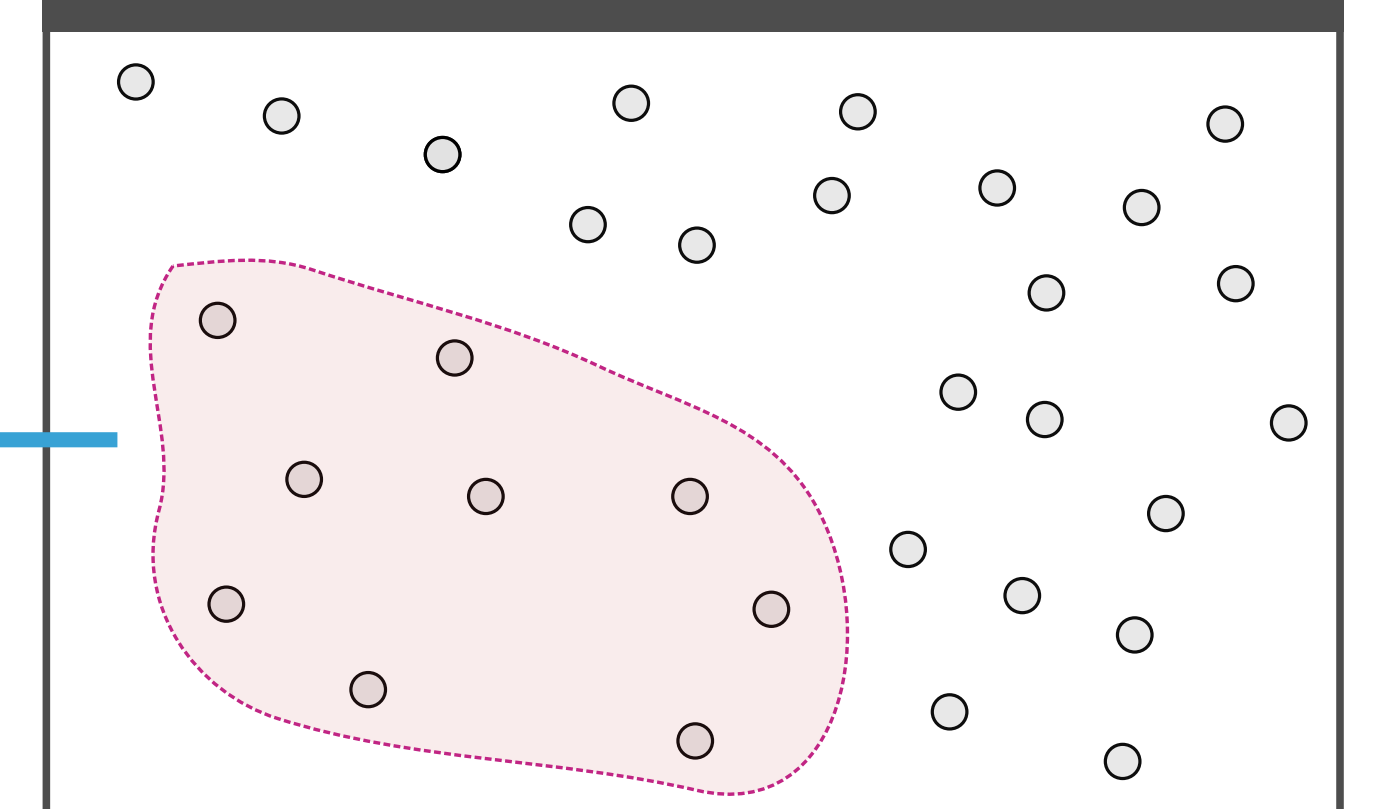


The subtrees of a hierarchical attribute are transformed into categorical attributes. The touring system compares each of these generated attributes with the given data subset.

Step 1: Selection

The users select a subset from the data they are currently analyzing as input for the touring system. The subset can be an attribute or an item set.

Items

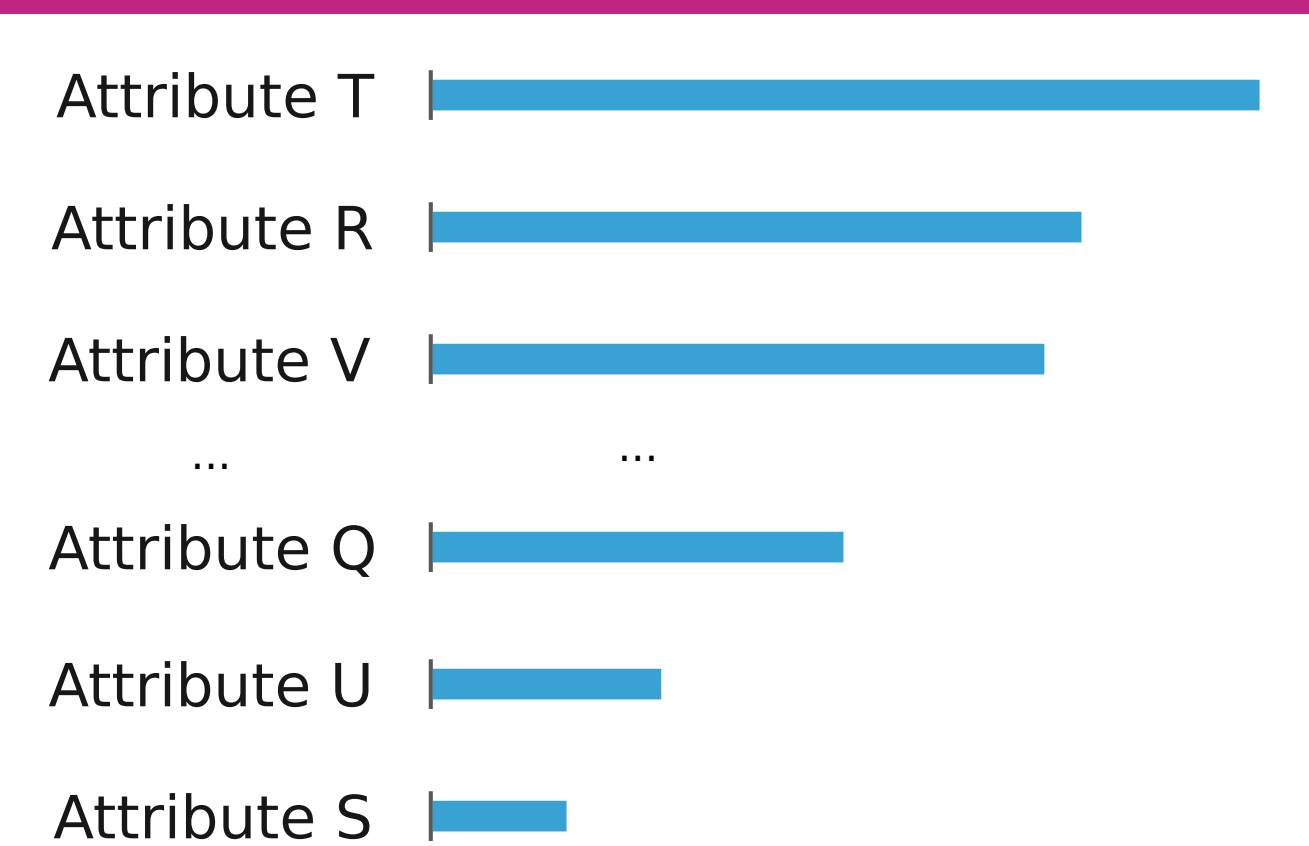


Step 2: Comparison

The selected subset is then compared with the whole dataset. For the comparison, all attributes become categorical by the discretization of numerical attributes and the categorization of hierarchical attributes. The touring system compares an attribute with all the other attributes and an item set is compared with each category of every attribute.

Step 3: Presentation

The similarity scores are presented as bars alongside their respective attributes. Users can include attributes with high similarity in the analysis, or vary the selected data subset and start the touring process again.



Results

We present a data-driven touring system for visual analytics tools that guides users in discovering relationships for a data subset of their interest. Based on the users' selection, attributes that show some kind of similarity are presented.

The described solution is generalized for data types and domains and independent of the employed visual analysis tool. We integrated the system in the visual analytics tools StratomeX [2] and Ordino [3].

Three heterogeneous multi-attribute datasets have been used to demonstrate data domain and data type generalization: Two datasets of the The Cancer Genome Atlas and a dataset of tumor patients, provided by the Kepler University Hospital Linz, Austria.

References

- [1] Ceneda, Davide, et al. "Characterizing guidance in visual analytics." IEEE Transactions on Visualization & Computer Graphics 1 (2017): 111-120.
- [2] Lex, Alexander, et al. "StratomeX: visual analysis of large-scale heterogeneous genomics data for cancer subtype characterization." Computer graphics forum. Vol. 31. No. 3pt3. Oxford, UK: Blackwell Publishing Ltd, 2012.
- [3] Streit, Marc, et al. "Ordino: visual analysis tool for ranking and exploring genes, cell lines, and tissue samples." bioRxiv (2018): 277848.

Acknowledgments

This work was done in collaboration with the Johannes Kepler University Linz and the Kepler University Hospital.

Contact: klaus.eckelt@jku.at